The Data Asset:

Databases, Business Intelligence, and Competitive Advantage

First draft version last modified: June 20, 2009 – comments welcome – john.gallaugher@bc.edu
Note: this is an earlier version of the chapter. All chapters updated after July 2009 are now hosted (and still free) at http://www.flatworldknowledge.com. For details see the 'Courseware' section of http://gallaugher.com

INTRODUCTION

The planet is awash in data. Cash registers ring up transactions worldwide. Web browsers leave a trail of cookie crumbs nearly everywhere they go. And with RFID, inventory can literally announce its presence so that firms can precisely journal every hop their products make along the value chain: "I'm arriving in the warehouse", "I'm on the store shelf", "I'm leaving out the front door".

A study by Gartner Research claims that the amount of data on corporate hard drives doubles every six months[1], while IDC states that the collective number of those bits already exceeds the number of stars in the universe[2]. Wal-Mart alone boasts a data volume nearly *30 times* as large as the *entire* print collection of the U.S. Library of Congress[3].

And with this flood of data comes a tidal wave of opportunity. Increasingly standardized corporate data, and access to rich, third party datasets; all leveraged by cheap, fast computing and easier-to-use software; are collectively enabling a new age of data-driven, fact-based decision making. You're less likely to hear old-school terms like decision support systems used to describe what's going on here. The phrase of the day is *business intelligence* (*BI*), a catchall term combining aspects of reporting, data exploration and ad-hoc queries, and sophisticated data modeling and analysis. Aside Business Intelligence in the new managerial lexicon is the phrase *analytics*, a term describing the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions[4].

The benefits of all this data and number crunching are very real, indeed. Data leverage lies at the center of competitive advantage we've studied in the Zara, Netflix, and Google cases. Data mastery has helped vault Wal-Mart to the top of the Fortune 500 list. It helped Harrah's Casino Hotels grow to be twice as profitable as similarly-sized Caesar's, and rich enough to acquire this rival. And data helped Capital One find valuable customers that competitors were ignoring, delivering ten-year financial performance a full 10 times greater than the S&P 500. Data-driven decision making is even credited with helping the Red Sox win their first World Series in 83 years, and with helping the New England Patriots win three Super Bowls in four years. To quote from the a BusinessWeek cover story on analytics, "Math will Rock Your World!"[5]

---

[1] Babcock, 2006.
[2] Mearian, 2008.
[3] Comparing Wal-Mart's 583 TB (Evans-Correia, 2006) to the Library of Congress estimate of 20 TB (Gewirtz, 2009).
[4] Davenport and Harris, 2007
[5] Baker, 2006.

Sounds great, but it can be a tough slog getting an organization to the point where it has a leveragable data asset. In many organizations data lies dormant, spread across inconsistent formats and incompatible systems, unable to be turned into anything of value. Many firms have been shocked at the amount of work and complexity required to pull together an infrastructure that empowers its managers. But not only can this be done, it must be done. Firms that are basing decisions on hunches aren't managing, they're gambling. And the days of uninformed managerial dice rolling are over.

While we'll study technology in this chapter, our focus isn't as much on the technology itself as it is on what you can do with that technology. Consumer products giant P&G believes in this distinction so thoroughly that the firm renamed its IT function as "Information and Decision Solutions"[6]. It's solutions that drive technology decisions, not the other way around.

In this chapter we'll study the data asset, how it's created, how it's stored, and how it's accessed and leveraged. We'll also study many of the firms mentioned above, and more; providing a context for understanding how managers are leveraging data to create winning models, and how those that have failed to realize the power of data have been left in the dust.

Data, Analytics, and Competitive Advantage

Anyone can acquire technology – but data is oftentimes considered a defensible source of competitive advantage. The data a firm can leverage is a true strategic asset when it's rare, valuable, imperfectly imitable, and lacking in substitutes (see the Strategy & Technology chapter).

If more data brings more accurate modeling, moving early to capture this rare asset can be the difference between a dominating firm and also ran. But be forewarned, there's no monopoly on math. Advantages based on capabilities and data that others can acquire will be short-lived. Those advances leveraged by the Red Sox were originally pioneered by the Oakland As, and are now used by nearly every team in the major leagues.

This doesn't mean that firms can ignore the importance data can play in lowering costs, increasing customer service, and other ways that boost performance. But differentiation will be key in distinguishing operationally effective data use from those efforts that can yield true strategic positioning.

DATA, INFORMATION, AND KNOWLEDGE

Data is simply raw facts and figures. Alone it tells you nothing. The real goal is to turn data into *information*. Data becomes information when it's presented in a context so that it can answer a question or support decision-making. And it's when this information can be combined with a manager's *knowledge* – their insight from experience and expertise – that stronger decisions can be made.

Trusting Your Data

The ability look critically at data and assess its validity are vital managerial skills. When decision-makers are presented with wrong data, the results can be disastrous. And these problems can get

[6] Soat, 2007.

amplified if bad data is fed to automated systems. As an example, look at the series of man-made and computer-triggered events that brought about a billion dollar collapse in United Airlines stock.

In the wee hours one Sunday morning in September, 2008, a single reader browsing back-stories on the Orlando Sentinel's website viewed a 2002 article on the bankruptcy of United Airlines (UAL went bankrupt in 2002, but emerged from bankruptcy four years later). That lone web-surfer's access of this story during such a low-traffic time was enough for the Sentinel's web server to briefly list the article as one of the paper's 'most popular'. Google crawled the site and picked up this 'popular' news item, feeding it into Google News.

Early that morning, a worker in a Florida investment firm came across the Google-fed story, assumed United had yet again filed for bankruptcy, then posted a summary on Bloomberg. Investors scanning Bloomberg jumped on what looked like a reputable early warning of another United bankruptcy, dumping UAL stock. Blame the computers again – the rapid plunge from these early trades caused automatic sell systems to kick in (event-triggered, computer-automated trading is responsible for about 30 percent of all stock trades). Once the machines took over, UAL dropped like a rock, falling from $12 to $3. That drop represented the vanishing of $1 billion in wealth, and all this because no one checked the date on a news story. Welcome to the new world of paying attention![7]

Understanding How Is Data Organized: Key Terms and Technologies

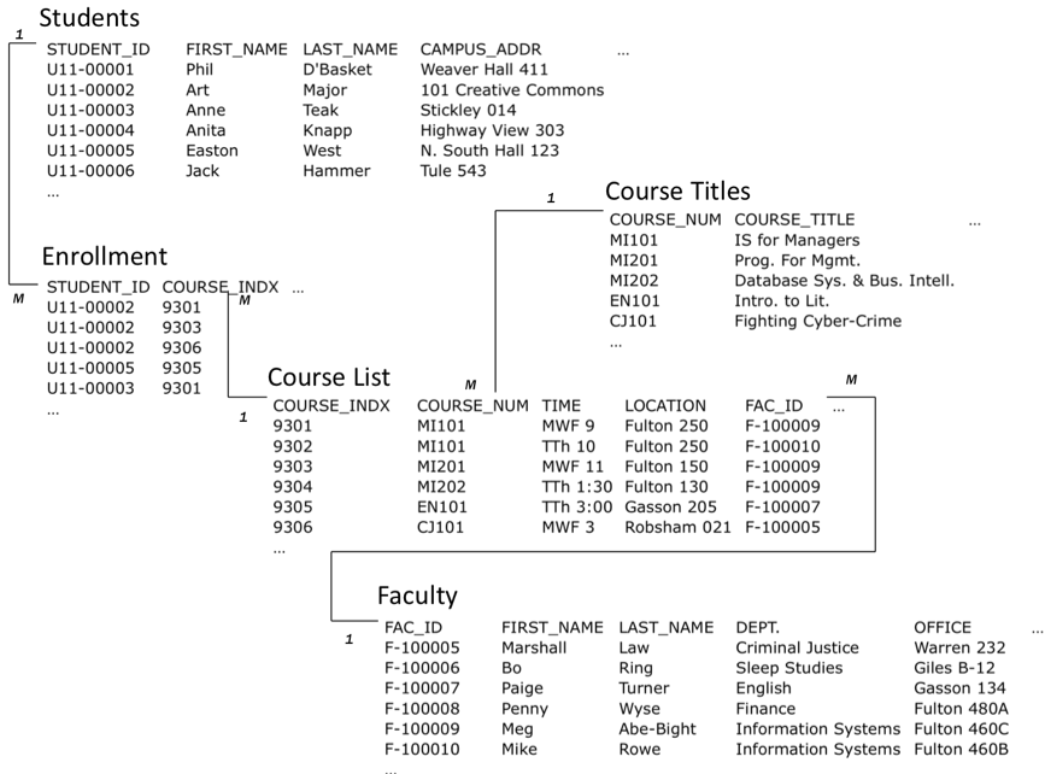
A *database* is simply a list or lists of information. Most organizations have several databases – perhaps even hundreds or thousands. And these various databases might be focused on any combination of functional areas (sales, product returns, inventory, payroll), geographical regions, or business units. Firms often create specialized databases for recording transactions, as well as databases that aggregate data from multiple sources in order to support reporting and analysis.

Databases are created and maintained using programs called *database management systems* (*DBMS*), sometimes referred to as *database software*. DBMS products vary widely in scale and capabilities. They include the single-user, desktop versions of Microsoft Access or Filemaker Pro, web-based offerings like Intuit QuickBase, and industrial strength products from Oracle, IBM (DB2), Sybase, Microsoft (SQL Server), and others. Oracle is the world's largest database software vendor, and database software has meant big bucks for Oracle co-founder and CEO Larry Ellison. Ellison perennially ranks in the Top 10 of the Forbes 400 list of wealthiest Americans.

The acronym SQL (often pronounced *sequel*) also shows up a lot when talking about databases. *SQL* or *structured query language*, is a language for creating and manipulating databases, and it is by far the most common database language. You'll find variants of SQL inhabiting everything from lowly desktop software, to high-powered enterprise products. Microsoft's high-end database is even called SQL Server. And of course there's also the open-source MySQL (acquired by Oracle as part of the firm's purchase of Sun Microsystems). Given this popularity, if you're going to learn one language for database use, SQL's a pretty good choice. And for a little inspiration, take a look on Monster.com, searching for jobs mentioning SQL. You'll likely see scores of listings, suggesting that while database systems have been good for Ellison, learning more about them might be pretty good for you, too.

[7] Harvey, 2008.

Even if you don't become a database programmer or *administrator*, you're almost surely going to be called upon to dive in and use a database. You may even be asked to help identify your firm's data requirements. It's quite common for non-tech employees to work on development teams with technical staff, defining business problems, outlining processes, setting requirements, and determining the kinds of data the firm will need to leverage. Database systems are powerful stuff, and can't be avoided, so a bit of understanding will serve you well.



A simplified relational database for a university course registration system

Here are some key concepts to help get you oriented:

- A *Table* or *file* refers to a list of data.
- A *database* is either a single table or a collection of related tables. The course registration database above depicts five tables.
- *Columns* or *fields* define the data that a Table can hold. The Students table above shows columns for STUDENT_ID, FIRST_NAME, LAST_NAME, CAMPUS_ADDR (the … symbols above are meant to indicate that in practice there may be more columns or rows than are shown in this simplified diagram).
- A *row* or *record* represents a single instance of whatever the table keeps track of. In the example above, each row of the Students table represents a student, each row of the Enrollment table represents the enrollment of a student in a particular course, and each row of the Course List represents a given section of each course offered by the University.
- A *key* is the field used to relate tables in a database. Look at how the STUDENT_ID key is used above. There's *one* unique STUDENT_ID for each student, but the STUDENT_ID may appear *many* times in the Enrollment table, indicating that each student may be enrolled

in many classes. The '1' and 'M' in the diagram above indicate the one to many relationships among the keys in these tables.

Databases organized like the one above, where multiple tables are related based on common keys, are referred to as *relational databases*. There are many other database formats (sporting names like *hierarchical*, and *object-oriented*), but relational databases are far and away the most popular. And all SQL databases are relational databases.

We've just scratched the surface for a very basic introduction. Expect that a formal class in database systems will offer you far more detail and better design principles than are conveyed in the elementary example above. But you're already well on your way!

Things to Think About Regarding the Example Above:

- What if you wanted to keep track of student majors? How would you do this? Would you modify an existing table? Would you add new tables? Why or why not?
- Why do you suppose we need a Course Title table?
- This database is simplified for our brief introduction. What additional data would you need to keep track of if this were a real course registration system? What changes would you make in the database above to account for these needs?

WHERE DOES DATA COME FROM?

Organizations can pull together data from a variety of sources. While the examples that follow aren't meant to be an encyclopedic listing of possibilities, they will give you a sense of the diversity of options available for data gathering.

Transaction Processing Systems

For most organizations that sell directly to their customers, *transaction processing systems* (*TPS*) represent a fountain of potentially insightful data. Every time a consumer uses a point-of-sale system, an ATM, or a service desk, there's a *transaction* (some kind of business exchange) occurring, and an event worth tracking.

But while TPS can generate a lot of bits, it's sometimes tough to match this data with a specific customer. For example, if you pay a retailer in cash, you're likely remain a mystery to your merchant because your name isn't attached to your money. Grocers and retailers can tie you to cash transactions if they can convince you to use a loyalty card. Use one of these cards and you're in effect giving up information about yourself in exchange for some kind of financial incentive. The explosion in retailer cards is directly related to each firm's desire to learn more about you and to turn you into a more loyal and satisfied customer.

Some cards provide an instant discount (think the CVS Pharmacy ExtraCare card), others allow you to build up points over time (Best Buy's Reward Zone). The latter has the additional benefit of acting as a switching cost. A customer may think "I could get the same thing at Target, but at Best Buy, it'll increase my existing points balance and soon I'll get a cash-back coupon".

Tesco: Tracked Transactions, Increased Insights, and Surging Sales

UK grocery giant Tesco, the planet's third largest retailer, is envied worldwide for what analysts say is the firm's unrivaled ability to collect vast amounts of retail data and translate this into sales[8].

Tesco's data collection relies heavily on its ClubCard loyalty program, an effort pioneered back in 1995. But Tesco isn't just a physical retailer. As the world's largest Internet grocer, the firm can mine both in-store purchases, as well as website visits. Remove products from your virtual shopping cart? Tesco can track this. Visited a product comparison page? Tesco watches which product you've chosen to go with, and which you've passed over. Done your research online, then traveled to a store to make a purchase? Tesco sees this, too.

Tesco then mines all this data to understand how consumers respond to factors such as product mix, pricing, and store layout. Consumer-level targeting allows the firm to tailor its marketing messages to specific subgroups, promoting the right offer through the right channel at the right time and the right price. To get a sense of Tesco's laser-focused targeting possibilities, consider that the firm sends out close to 10 million different, targeted offers each quarter.[9] Offer redemption rates are the best in the industry, with some coupons scoring an unheard of 90 percent usage[10].

The firm's data driven management is clearly delivering results. In April 2009, while operating in the teeth of a global recession, Tesco posted record corporate profits and the highest earnings ever for a British retailer[11].

Enterprise Software (CRM, SCM, and ERP)

Firms increasingly set up systems to gather additional data beyond conventional purchase transactions or website monitoring. CRM or customer –relationship management systems are often used to empower employees to track and record data at nearly every point of customer contact. Someone calls for a quote? Brings a return back to a store? Writes a complaint e-mail? A well-designed CRM system can capture all these events for subsequent analysis or for triggering follow-up events.

Enterprise software includes not just CRM systems, as but also categories that touch every aspect of the value chain, including supply chain management (SCM) and enterprise resource planning (ERP) systems. More importantly, enterprise software tends to be more integrated and standardized than the prior era of proprietary systems that many firms developed themselves. This helps in combining data across business units and functions, and in getting that data into a form where it can be turned into information (for more on enterprise systems, see the chapter *Software: A Primer*).

Surveys

Sometimes firms supplement operational data with additional input from surveys and focus groups. Oftentimes direct surveys can tell you what your cash register can't. Zara store

[8] Capell, 2008.
[9] Davenport and Harris, 2007b.
[10] Lowenstein, 2002.
[11] Capell, 2009.

managers informally survey customers in order to help shape designs and product mix. Online grocer FreshDirect (see the Strategy and Technology chapter) surveys customers weekly, and has used this feedback to drive initiatives from reducing packaging size to including star ratings on produce[12]. Many CRM products also have survey capabilities that allow for additional data gathering at all points of customer contact.

Can Technology 'Cure' U.S. Healthcare?

The American healthcare system is broken. It's costly, inefficient, and problems seem to be getting worse. Estimates suggest that healthcare spending makes up a whopping 18 percent of U.S. gross domestic product[13]. U.S. Automakers spend more on healthcare than they do on steel[14]. Even more disturbing, it's believed that medical errors cause as many as 98,000 unnecessary deaths in the U.S. each year; more than motor vehicle accidents, breast cancer, or AIDS[15].

For years it's been claimed that technology has the potential to reduce errors, improve health care quality, and save costs. Now pioneering hospital networks and technology companies are partnering to help tackle cost / quality issues. For a look at possibilities for leveraging data throughout the doctor-patient value chain, consider the "event-driven medicine" system built by Dr. John Halamka and his team at Boston's Beth Israel Deaconess Medical Center (part of the Harvard Medical School network).

When docs using Halamka's system encounter a patient with a chronic disease, they generate a decision support 'screening sheet'. Each event in the system: an office visit, a lab results report (think the medical equivalent of transactions and customer interactions), updates the patient database. Combine that electronic medical record information with *artificial intelligence* on best practice, and the system can offer recommendations for care, such as "patient is past due for an eye exam" or "patient should receive pneumovax [a vaccine against infection] this season."[16] The systems don't replace decision-making by doctors and nurses, but it does help to ensure that key issues are on a provider's radar.

More efficiencies and error checks show up when prescribing drugs. Docs are presented with a list of medications covered by that patient's insurance, allowing them to choose quality options while controlling costs. Safety issues, guidelines, and best practices are also displayed. When correct, safe medication in the right dose is selected, it is routed to the patients' pharmacy of choice, as Halamka puts it, going from "doctor's brain to patients vein" without any of that messy physician handwriting, all while squeezing out layers where errors from human interpretation or data entry might occur.

President Obama believes technology initiatives can save healthcare as much as $120 billion a year, or roughly $2,500 per family[17]. If systems like Halamka's realize their promise, big benefits may be just around the corner.

External Sources

Sometimes it makes sense to combine a firm's data with bits brought in from the outside. Many firms, for example, don't sell directly to consumers (this includes most drug companies and

[12] Braddock, 2009.
[13] Zhang, 2009.
[14] Milligan, 2009.
[15] Appleton, 2009; Obama, 2009.
[16] Halamka, 2009.
[17] McCullagh, 2009.

packaged goods firms). If someone else sells your products for you, then you'll likely rely heavily on outside data.

Data bought from sources available to all might not yield competitive advantage on its own, but it can provide key operational insight for increased efficiency and cost savings. And when combined with a firm's unique data assets, it may give firms a high-impact edge.

Restaurant chain Brinker runs 1700 eateries in 27 countries under the Chili's, On The Border, and Maggiano's brands. The firm (whose ticker symbol is EAT), supplements their own data with external feeds on weather, employment statistics, gas prices, and other factors, and uses this in predictive models that help the firm in everything from determining staffing levels to switching around menu items[18].

Carnival Cruise Lines combines its own customer data with third-party information tracking household income and other key measures. This data plays a key role in a recession, since it helps the firm target limited marketing dollars on those past customers that are more likely to be able to afford to go on a cruise. So far it's been a winning approach. For three years in a row, the firm has experienced double-digit increases in bookings by repeat customers[19].

Who's Collecting Data About You?

There's a thriving industry collecting data about you. Buy from a catalog, fill out a warrantee card, or have a baby, and there's a very good chance that this event will be recorded in a database somewhere, added to a growing digital dossier that's made available for sale to others. If you've ever gotten catalogs, coupons, or special offers from firms you've never dealt with before, this was almost certainly a direct result of a behind-the-scenes trafficking in the 'digital you'.

Firms that trawl for data and package it up for resale are known as *data aggregators*. They include Acxiom, a $1.3 billion a year business that combines public source data on real estate, criminal records, and census reports with private information from credit card applications, warranty card surveys, and magazine subscriptions. The firm holds data profiling some 200 million Americans[20].

Or maybe you've heard of Lexis-Nexis. Many large universities subscribe to the firm's electronic newspaper, journal, and magazine databases. But the firm's parent, Reed Elsevier, is a data sales giant, with divisions packaging criminal records, housing information, and additional data used to uncover corporate fraud and other risks. In Feb. 2008 the firm got even more data-rich, acquiring Acxiom competitor ChoicePoint for $4.1 billion. With that kind of money involved, it's clear that data aggregation is very big business[21].

The Internet also allows for easy access to data that had been public, but otherwise difficult to access. For one example, consider home sale prices and home value assessments. While technically in the public record, someone wanting this information previously had to traipse down to their Town Hall and speak to a clerk, who would hand over a printed log book. Not exactly a Google-speed query. Contrast this with a visit to Zillow.com. The free site lets you pull up a map of your town and instantly peek at how much your neighbors paid for their homes. And it lets them see how much you paid for yours, too.

[18] King, 2009.
[19] King, 2009.
[20] Gefter and Simonite, 2008.
[21] Greenberg, 2008.

Computerworld's Robert Mitchell uncovered a more disturbing issue when public record information is made available online. His New Hampshire municipality had digitized and made available some of his old public documents without obscuring that holy grail for identity thieves, his social security number[22].

Then there are accuracy concerns. A record incorrectly identifying you as a cat lover is one thing, but being incorrectly named to the terrorist watch list is quite another. During a five week period airline agents tried to block one U.S. citizen from boarding airplanes on five separate occasions simply because his name resembled an alias used by a suspected terrorist. That citizen? Massachusetts Senator Ted Kennedy[23].

For the data trade to continue, firms will have to treat customer data as the sacred asset it is. Step over that 'creep-out' line, and customers will push back, increasingly pressing for tighter privacy laws. Data aggregator Intellius used to track cell phone customers, but backed off in the face of customer outrage and threatened legislation.

Another concern: sometimes data aggregators are just plain sloppy, committing errors that can be costly for the firm and potentially devastating for victimized users. Examples: from 2002 to 2003, a hacker stole 1.6 billion records from Acxiom; while in 2005, ChoicePoint accidentally sold records on 145,000 individuals to a cybercrime identity theft ring. The ChoicePoint case resulted in a $15 million fine from the Federal Trade Commission[24]. Just because you can gather data and traffic in bits doesn't mean that you should. Any data-centric effort should involve input not only from business and technical staff, but from the firm's legal team, as well (for more, see box *Privacy Regulation, A Moving Target*).

DATA RICH, INFORMATION POOR

Despite being awash in data, many organizations are data rich but information poor. A survey by consulting firm Accenture found 57 percent of companies reporting that they didn't have a beneficial, consistently updated, companywide analytical capability. Among major decisions, only 60 percent were backed by analytics. 40 percent were made by intuition and gut instinct[25]. The big culprit limiting BI initiatives is getting data into a form where it can be used – analyzed and turned into information. Here's a look at some factors holding back information advantages, and how firms are preparing data for analytics.

Incompatible Systems

There are many reasons for an info drought amidst a data deluge. Just because data is collected doesn't mean it can be used. This is a big problem for large firms that have *legacy systems*, outdated information systems that were not designed to share data, aren't compatible with newer technologies, and aren't aligned with the firm's current business needs. The problem can be made worse by mergers and acquisitions, especially if a firm depends on operational systems that are incompatible with its partner. Merging firms might also be under extended contract with different vendors or outsourcers, further complicating attempts to combine technology resources. Folks working in *M&A* (the area of investment banking focused on valuing and facilitating

[22] Mithcell, 2009.
[23] Swarns, 2004.
[24] Greenberg, 2008.
[25] King, 2009.

mergers and acquisitions) beware – it's critical to uncover these hidden costs of technology integration before deciding if a deal makes financial sense.

Legacy Systems: A Prison for Strategic Assets

The experience of one Fortune 100 firm that your author has worked with illustrates how incompatible information systems can actually hold back strategy. This firm was the largest in its category, and sold identical commodity products sourced from its many plants worldwide. Being the biggest should have given the firm scale advantages. But information systems used by the firm's separate manufacturing facilities were developed before worldwide networking was commonplace. Still more plants were acquired through acquisition, each coming with its own separate and incompatible legacy systems.

The plants with different information systems used *different* part numbers and naming conventions even though they sold *identical* products. As a result, the firm had no timely information on how much of a particular item was sold to which worldwide customers. The company was essentially operating as a collection of smaller, regional businesses, rather than as the worldwide behemoth that it was.

After the firm developed an information system that standardized data across these plants, it was able to get a single view on worldwide sales. The firm then used this data to approach their biggest customers, negotiating lower prices in exchange for increased commitments in worldwide purchasing. This let the firm take share from regional rivals. It also gave the firm the ability to shift manufacturing capacity globally, as currency prices, labor conditions, disaster, and other factors impacted sourcing. The new information system in effect liberated the latent strategic asset of scale, increasing sales by well over a billion and a half dollars in the four years following implementation.

Operational Data Can't Always Be Queried

Another problem turning data into information: most transactional databases aren't set up to be simultaneously accessed for reporting and analysis. When a customer buys something from a cash register, that action may post a sales record, and deduct an item from the firm's inventory. So any requests made to the database is performed pretty quickly – the system adds or modifies the few records involved and it's done – in and out in a flash.

But if a manager asks a database to analyze historic sales trends showing the most and least profitable products, they may be asking a computer to look at thousands of transaction records, comparing results, and neatly ordering findings. That's not a quick in and out task, and it may very well require significant processing to come up with the request. Do this against the very databases you're using to record your transactions, and you might grind your computers to a halt.

Getting data into systems that can support analytics is where Data Warehouses and Data Marts come in.

Privacy Regulation – A Moving Target

New methods for tracking and gathering user information appear daily, testing user comfort levels. For example, the firm Umbria uses software to analyze millions of blog and forum posts every day, using sentence structure, word choice, and quirks in punctuation to determine a blogger's gender, age, interests, and opinions. In 2009, Apple introduced facial recognition software while integrating iPhoto into Facebook. It's quite possible that in the future, someone will be able to upload a photo to a service and direct it to find all the accessible photos and video on the Internet that match that person's features.

And while targeting is getting easier, a Carnegie Mellon study showed that it doesn't take much to find someone with a minimum of data. Simply by knowing gender, birth date, and postal zip code, 87 percent of people in the U.S. could be pinpointed by name[26].

Some feel that Moore's Law, the falling cost of storage, and the increasing reach of the Internet have us on the cusp of a privacy train wreck. And that may inevitably lead to more legislation that restricts data-use possibilities. Noting this, strategists and technologists need to be fully aware of the legal environment their systems face (see Google chapter for examples and discussion) and consider how such environments may change in the future. Many industries have strict guidelines on what kind of information can be collected and shared.

For example, HIPAA (the U.S. Health Insurance Portability and Accountability Act) includes provisions governing data use and privacy among healthcare providers, insurers, and employers. The Financial industry has strict requirements for recording and sharing communications between firm and client (among many other restrictions). There are laws limiting the kinds of information that can be gathered on younger web surfers. And there are several laws operating at the state level, as well.

International laws also differ from those in the U.S. Europe, in particular, has a strict European Privacy Directive, governing provisions that limits data collection, requires notice and approval of many types of data collection, and requires firms to make data available to customers with provisions for stopping collection efforts and correcting inaccuracies. Data-dependent efforts plotted for one region may not fully translate in another effort if the law limits key components of technology use. The constantly changing legal landscape also means that what works today might not be allowed in the future.

Firms beware – the public will almost certain demand tighter controls if the industry is perceived as behaving recklessly or inappropriately with customer data.

DATA WAREHOUSES AND DATA MARTS

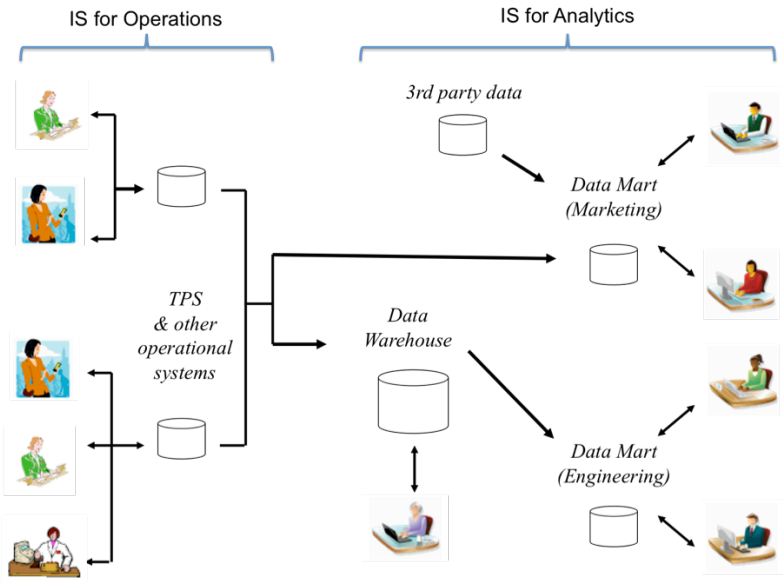
Since running analytics against transactional data can bog down a system, and since most organizations need to combine and re-format data from multiple sources, firms typically need to create separate data repositories for their reporting and analytics work – a kind of staging area from which to turn that data into information.

Two terms you'll hear for these kinds of repositories are *data warehouse* and *data mart*. A *data warehouse* is a set of databases designed to support decision-making in an organization. It is structured for fast online queries and exploration. Data warehouses may aggregate enormous amounts of data from many different operational systems.

A *data mart* is a database focused on addressing the concerns of a specific problem (e.g. increasing customer retention, improving product quality) or business unit (e.g. Marketing, Engineering).

Marts and warehouses may contain huge volumes of data. For example, a firm may not need to keep large amounts of point-of-sales or transaction data in its operational systems, but it might want historical data in its data mart because managers will be looking to see if they can discover relevant patterns that occur over time.

[26] Gefter and Simonite, 2008.



Information systems supporting operations (such as TPS) are typically separate, and 'feed' information systems used for analytics (such as data warehouses and data marts)

It's easy for firms to get seduced by a software vendor's demonstration showing data at your fingertips, presented in pretty graphs. But as mentioned earlier, getting data in a format that can be used for analytics is hard, complex, and challenging work. Large data warehouses can cost millions and take years to build. Every dollar may lead to five to seven more dollars on consulting and other services[27].

Most firms will face a tradeoff – do we attempt a large scale integration of the whole firm, or more targeted efforts with quicker payoffs? Firms in fast-moving industries or with particularly complex businesses may struggle to get sweeping projects completed in enough time to reap benefits before business conditions change. Most consultants now advise smaller projects with narrow scope driven by specific business goals[28].

Firms can eventually get to a unified data warehouse, but it may take time. Even analytics king Wal-Mart is just getting to that point. In 2007, it was reported that Wal-Mart had 700 different data marts, and hired Hewlett Packard for help in bringing the systems together to form a more integrated data warehouse[29].

The old saying from the movie Field of Dreams "If you build it, they will come" doesn't hold up well for large-scale data analytics projects. This work should start with a clear vision with business-focused objectives. When senior executives can see objectives illustrated in potential payoff, they'll be able to champion the effort, and experts agree, having an executive champion

[27] King, 2009.
[28] Rigby and Ledignham, 2004; King, 2009b
[29] Havenstein, 2007.

is a key success factor. Focusing on business issues will also drive technology choice, with the firm better able to focus on products that best fit its needs.

Once a firm has business goals and hoped-for payoffs clearly defined, it can address the broader issues needed to design, develop, deploy, and maintain its system. Issues include:[30]

- *Data relevance* – What data is needed to compete on analytics and to meet our current and future goals?
- *Data sourcing* – Can we even get the data we'll need? Where can this data be obtained from? Is it available via our internal systems? Via third party data aggregators? Via suppliers or sales partners? Do we need to set up new systems, surveys, and other collection efforts to acquire the data we need?
- *Data quantity* – How much data is needed?
- *Data quality* – Can our data be trusted as accurate? Is it clean, complete, and reasonably free of errors? How can the data be made more accurate and valuable for analysis? Will we need to 'scrub', calculate, and consolidate data so that it can be used?
- *Data hosting* – Where will the systems be housed? What are the hardware and networking requirements for the effort?
- *Data governance* – What rules and processes are needed to manage data from its creation through its retirement? Are there operational issues (backup, disaster recovery)? Legal issues? Privacy issues? How should the firm handle security and access?

Addressing all of these issues is beyond the scope of our managerial discussion, but let's provide a brief example of the kind of complexity a large firm might face in pulling together a data warehouse or data mart.

An executive from one of the largest banks in the U.S. stopped by my class to talk about the firm's data warehousing effort and lamented at how difficult it was to get his systems to do something as simple as properly distinguishing between men and women. The company's customer-focused data warehouse drew data from 36 separate operational systems – bank teller systems, ATMs, student loan reporting systems, car loan systems, mortgage loan systems, and more. Collectively these legacy systems expressed gender in *seventeen* different ways: "M" or "F"; "m" or "f"; "Male" or "Female"; "MALE" or "FEMALE", 1 for man, 0 for woman; 0 for man, 1 for woman; and more, plus various codes for 'unknown'. The best math in the world is of no help if the values used aren't any good. There's a saying in the industry is *garbage in, garbage out*.

e-Discovery: Supporting Legal Inquiries

Data archiving isn't just for analytics. Sometimes the law requires organizations to dive into their electronic records. *E-Discovery* refers to identifying and retrieving relevant electronic information to support litigation efforts. E-Discovery is something a firm should account for in its archiving and data storage plans. Unlike analytics that promise a boost to the bottom line, there's no profit in complying with a judge's order – it's just a sunk cost. But organizations can be compelled by court order to scavenge their bits, and the cost to uncover difficult to access data can be significant, if not planned for in advance.

[30] Key points adapted from Davenport and Harris, 2009.

In one recent example, the Office of Federal Housing Enterprise Oversight (OFHEO) was subpoenaed for documents in litigation involving mortgage firms Fannie Mae and Freddie Mac. Even though the OFHEO wasn't a party in the lawsuit, the agency had to comply with the search – an effort that cost $6 million dollars, a full 9% of its total yearly budget[31].

THE BUSINESS INTELLIGENCE TOOLKIT

So far we've discussed where data can come from, and how we can get data into a form where we can use it. But how, exactly, do firms turn that data into information? That's where the various software tools of Business Intelligence and Analytics come in. Potential products in the Business Intelligence toolkit range from simple spreadsheets to ultra-sophisticated data mining packages leveraged by teams employing 'rocket-science' mathematics.

Query and Reporting Tools

The idea behind query and reporting tools is to present users with a subset of requested data, selected, sorted, ordered, calculated, and compared, as needed. Managers use these tools to see and explore what's happening inside their organizations.

Canned reports provide regular summaries of information in a pre-determined format. *Ad-hoc reports* allow users to dive in, selecting fields, ranges, and other parameters to build their own reports on the fly. *Dashboards* provide a sort of heads-up display of critical indicators, letting managers get a graphical glance at key performance metrics. Some tools may allow data to be exported into spreadsheets. Yes, even the lowly spreadsheet can be a powerful tools for modeling 'what if' scenarios and creating additional reports (of course be careful, if data can be exported then it can potentially leave the firm inappropriately, raising privacy, security, legal, and competitive concerns).

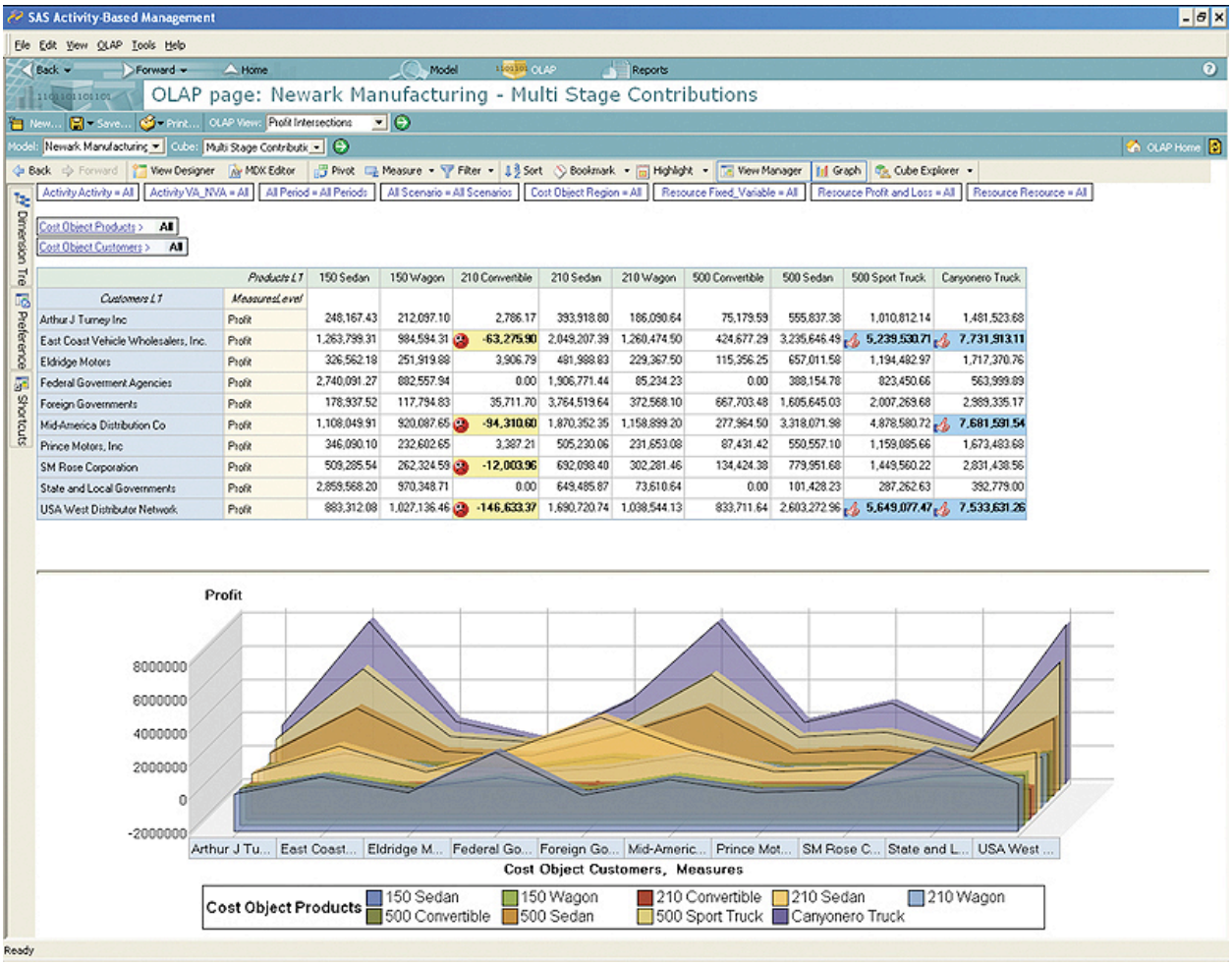
[31] Conry-Murray, 2009.

A business intelligence dashboard[32]

A sub-category of reporting tools is referred to as *OLAP* (pronounced 'oh-lap'), for *online analytical processing*. Data used in OLAP reporting is usually sourced from standard relational databases, but it's calculated and summarized in advance, across multiple dimensions, with the data stored in a a special database called a *data cube*. This extra setup step makes OLAP fast (sometimes 1000 times faster than performing comparable queries against conventional relational databases).

A manager using an OLAP tool to quickly explore and compare data across multiple factors such as time, geography, product lines, etc. In fact, OLAP users often talk about how they can 'slice and dice' their data, 'drilling down' inside the data to uncover new insights. And while conventional reports are usually presented as summarized list of information, OLAP results look more like a spreadsheet, with the various dimensions of analysis in rows and columns, and summary values at the intersection of these columns.

[32] Note to Editor: image from http://www.oracle.com/technology/pub/articles/rittman_dash.html. Obtain permission for commercial use, or find a similar image that can be used.



Customers L1	Measures Level	Products L1 150 Sedan	150 Wagon	210 Convertible	210 Sedan	210 Wagon	500 Convertible	500 Sedan	500 Sport Truck	Canyonero Truck
Arthur J Turney Inc	Profit	248,167.43	212,097.10	2,786.17	393,918.80	186,090.64	75,179.59	555,837.38	1,010,812.14	1,481,523.68
East Coast Vehicle Wholesalers, Inc.	Profit	1,263,799.31	984,594.31	-63,275.90	2,049,207.39	1,260,474.50	424,677.29	3,235,646.49	5,239,530.71	7,731,913.11
Eldridge Motors	Profit	326,562.18	251,919.88	3,906.79	481,988.83	229,367.50	115,356.25	657,011.58	1,194,482.97	1,717,370.76
Federal Goverment Agencies	Profit	2,740,091.27	882,557.94	0.00	1,906,771.44	85,234.23	0.00	388,154.78	823,450.66	563,999.89
Foreign Governments	Profit	178,937.52	117,794.83	35,711.70	3,764,519.64	372,568.10	667,703.48	1,605,645.03	2,007,269.68	2,989,335.17
Mid-America Distribution Co	Profit	1,108,049.91	920,087.65	-94,310.60	1,870,352.35	1,158,899.20	277,964.50	3,318,071.98	4,878,580.72	7,681,591.54
Prince Motors, Inc	Profit	346,090.10	232,602.65	3,387.21	505,230.06	231,653.08	87,431.42	550,557.10	1,159,085.66	1,673,483.68
SM Rose Corporation	Profit	509,285.54	262,324.59	-12,003.96	692,098.40	302,281.46	134,424.38	779,951.68	1,449,560.22	2,831,438.56
State and Local Governments	Profit	2,859,568.20	970,348.71	0.00	649,485.87	73,610.64	0.00	101,428.23	287,262.63	392,779.00
USA West Distributor Network	Profit	883,312.08	1,027,136.46	-146,633.37	1,690,720.74	1,038,544.13	833,711.64	2,603,272.96	5,649,077.47	7,533,631.26

Image from an OLAP report comparing multiple dimensions. Company is along the vertical axis, product is along the horizontal access. Many OLAP tools can also present graphs of multi-dimensional data[33]

Public Sector Reporting Tools in Action: Fighting Crime and Fighting Waste

Access to ad-hoc query and reporting tools can empower all sorts of workers. Consider what analytics tools have done for the Richmond, VA police force. The city provides department investigators with access to data from internal sources such as 911 logs and police reports, and combines this with outside data including neighborhood demographics, payday schedules, weather reports, traffic patterns, sports events, and more.

Experienced officers dive into this data, exploring when and where crimes occur. These insights help the department decide how to allocate its limited policing assets to achieve the biggest impact. While IT staffers put the system together, the tools are actually used by officers with expertise in fighting street crime – the kinds of users with the knowledge to hunt down trends and interpret the causes behind the data. And it seems this data helps make smart cops even smarter - the system is credited with delivering a single year crime-rate reduction of 20 percent[34].

[33] Editor Note: SAS image from: http://www.sas.com/solutions/abm/. Obtain permission for commercial use, or find a similar image that can be used.
[34] Lohr, 2008.

As it turns out, what works for cops also works for bureaucrats. When administrators for the City of Albuquerque were given access to ad-hoc reporting systems, they uncovered all sorts of anomalies, prompting excess spending cuts on everything from cell phone usage to unnecessarily scheduled overtime. And once again, BI performed for the public sector. The Albuquerque system delivered the equivalent of $2 million in savings in just the first three weeks it was used[35].

Data mining

While reporting tools can help users explore data, modern datasets can be so large that it might be impossible for humans to spot underlying trends. That's where data mining can help. *Data mining* is the process of using computers to identify hidden patterns and to build models from large data sets.

Some of the key areas where businesses are leveraging data mining include:

- customer segmentation - figuring out which customers are likely to be the most valuable to a firm.
- marketing and promotion targeting – identifying which customers will respond to which offers at which price at what time.
- market basket analysis – determining which products customers buy together, and how an organization can use this information to cross-sell more products or services.
- collaborative filtering – personalizing an individual customer's experience based on the trends and preferences identified across similar customers.
- customer churn – determining which customers are likely to leave, and what tactics can help the firm avoid unwanted defections.
- fraud detection – uncovering patterns consistent with criminal activity.
- financial modeling – building trading systems to capitalize on historical trends.
- hiring and promotion – identifying characteristics consistent with employee success in the firm's various roles.

For data mining to work, two critical conditions need to be present. 1) the organization must have clean, consistent data, and 2) the events in that data should reflect current and future trends. The recent financial crisis provides lessons on what can happen when either of these conditions isn't met.

First lets look at problems with using bad data: A report in the *New York Times* has suggested that in the period leading up to the 2008 financial crisis, some banking executives deliberately deceived risk management systems in order to skew capital-on-hand requirements. This let firms load up on risky debt, while carrying less cash for covering losses[36]. Deceive your systems with bad data and your models are worthless. In this case, wrong estimates from bad data left firms grossly overexposed to risk. When debt defaults occurred; several banks failed, and we entered the worst financial crisis since the Great Depression.

Now consider the problem of historical consistency: Computer-driven investment models can be very effective when the market behaves as it has in the past. But models are blind when faced

[35] Mulcahy, 2007.
[36] Hansell, 2008.

with the equivalent of the "100 year flood" – an anomaly or condition so extreme and unusual that it never showed up in the data used to build the model.

We saw this in the late 90s with the collapse of the investment firm Long Term Capital Management. LTCM was started by Nobel Prize-winning economists, but when an unexpected Russian debt crisis caused the markets to move in ways not anticipated by its models, the firm lost 90 percent of its value in less than two months. The problem was so bad that the Fed had to step in to supervise the firm's multi-billion dollar bailout. Fast forward a decade to the banking collapse of 2008, and we again see computer-driven trading funds plummet in the face of another unexpected event - the burst of the housing bubble.[37]

Data mining presents a host of other perils, as well. It's possible to over-engineer a model, building it with so many variables that the solution arrived at might only work on the subset of data you've used to create it. You might also be looking at a random but meaningless statistical fluke. In demonstrating how flukes occur, one quant investment manager uncovered a particularly strong predictor for historical prices in the S&P500 stock index. That predictor? Butter production in Bangladesh[38]. Sometimes durable and useful patterns just aren't in your data.

One way to test to see if you're looking at a random occurrence in the numbers is to divide your data, building your model with one portion of the data, and using another portion to verify your results. This is the approach Netflix has used to test results achieved by teams in the Netflix Prize, the firm's million-dollar contest for improving the predictive accuracy of its movie recommendation engine (see the *Netflix* case).

Finally, sometimes a pattern is uncovered, but determining the best choice for a response is less clear. As an example, let's return to the data-mining wizards at Tesco. An analysis of product sales data showed several money-losing products, including a type of bread known as 'milk loaf'. Drop those products, right? Not so fast. Further analysis showed milk loaf was a 'destination product' for a loyal group of high-value customers, and that these customers would shop elsewhere if milk loaf disappeared from Tesco shelves. The firm kept the bread as a loss-leader and retained those valuable milk loaf fans[39]. Data miner beware – first findings don't always reveal an optimal course of action.

This last example underscores the importance of recruiting a data mining and business analytics team that possesses three critical skills: information technology (for understanding how to pull together data, and for selecting analysis tools), statistics (for interpreting the strength and validity of results), and business knowledge (for helping set system goals, requirements, and offering deeper insight into what the data really says about the firm's operating environment). Miss one of these key functions, and your team could make some major mistakes.

---

[37] Wahba, 2008.
[38] Coy, 1997.
[39] Helm, 2008.

While we've focused on tools in our discussion above, many experts suggest that business intelligence is really an organizational processes as much as it is a set of technologies. Having the right team is critical in moving the firm from goal setting through execution and results.

Artificial Intelligence

Data Mining has its roots in a branch of computer science known as *artificial intelligence* (or *AI*). The goal of AI is create computer programs that are able to mimic or improve upon functions of the human brain. Data mining can leverage *neural networks* or other advanced algorithms and statistical techniques to hunt down and expose patterns, and build models to exploit findings.

Expert systems are AI systems that leverage rules or examples to perform a task in a way that mimics applied human expertise. Expert systems are used in tasks ranging from medical diagnoses to product configuration.

Genetic algorithms are model building techniques where computers examine many potential solutions to a problem, iteratively modifying (mutating) various mathematical models, and comparing the mutated models to search for a best alternative. Genetic algorithms have been used to build everything from financial trading models to handling complex airport scheduling, to designing parts for the international space station[40].

While AI is not a single technology, and not directly related to data creation, various forms of AI can show up as part of analytics products, CRM tools, transaction processing systems, and other information systems.

DATA ASSET IN ACTION: PRODUCT AND SERVICE-SECTOR EXAMPLES

In order to illustrate how firms create and leverage their data asset, we present two cases:

- Wal-Mart demonstrates how a physical product retailer can create and leverage a data asset to achieve world-class supply chain efficiencies targeted primarily at driving down costs.
- Harrah's Entertainment provides an example of exceptional data asset leverage in the service sector, focusing on how this technology enables world-class service through customer relationship management.

Both of these cases discuss technology use, but also end with the substantial challenges each firms face. As in all of our cases, we realize that technology is not a panacea, and no competitive advantage lasts forever.

TECHNOLOGY AND THE RISE OF WAL-MART

Wal-Mart is the world's number one retailer. The firm is so big that in three months it sells more than the number two U.S. retailer, Home Depot, sells in an entire year. Wal-Mart isn't just the largest retailer in the world, for several years now it has popped in and out of the top spot on the Fortune 500 list – meaning it has had revenues greater than *any* firm in the United States[41].

[40] Adapted from Kahn, 2002; Port, 2002; and McKay, 2009.
[41] From 2006 through 2009, Wal-Mart has appeared as either number one or number two in the Fortune 100 rankings.

At that size, it's clear that Wal-Mart's key source of competitive advantage is scale. But firms don't get scale overnight. Wal-Mart grew by leveraging information technology to an extent never before seen in the retail industry. To get a sense of the firm's tech-fueled efficiencies, a McKinsey study has suggested that 12 percent of U.S. economic productivity gains at the close of the prior decade were attributable to Wal-Mart[42]. The firm's tech prowess is so respected that many senior Wal-Mart IT executives have been snatched up for top roles at Dell, HP, Amazon, and Microsoft. The firm developed its tech prowess even though it's located far from the U.S. technology hubs of Silicon Valley, Boston, and Seattle. Wal-Mart is headquartered in Bentonville, Arkansas.

A Data Driven Value Chain

The Wal-Mart efficiency dance starts with a proprietary system called Retail Link. Developed in 1991 and continually refined ever since, Retail Link coordinates Wal-Mart's value chain from head to tail. Each time an item is scanned by a Wal-Mart cash register, Retail Link not only records the sale, but triggers inventory reordering and scheduling. This keeps shelves stocked, while keeping inventories at a minimum. An AMR report ranked Wal-Mart as having the 7th best supply chain in the country (the only other retailer in the top 20 was Tesco, at number 15)[43]. The firm's annual *inventory turnover ratio* of 8.4, means that Wal-Mart sells the equivalent of its entire inventory roughly every six weeks.

Back-office scanners keep track of inventory as it comes in. Suppliers are tracked based on parameters such as sales figures, return rates, and timeliness of deliveries. Suppliers have to be quick. When Levis' joined Wal-Mart in 2002 they had to guarantee they could replenish shelves in just 2 days. No other retailer had ever pushed it beyond five days. In order to avoid a tractor-trailer traffic jam in the Wal-Mart parking lot, deliveries are choreographed to arrive at intervals less than 10 minutes apart[44].

Wal-Mart has been a catalyst for technology adoption among its suppliers. The firm is currently leading an adoption effort that requires partners to leverage RFID to track and coordinate inventories. While the rollout has been slow, a recent P&G trial showed RFID boosted sales nearly 20 percent by ensuring that inventory was on shelves and located where it should be[45].

Data Mining Prowess

Wal-Mart also mines its mother lode of data to get its product mix right under all sorts of varying environmental conditions, protecting the firm from "a retailer's twin nightmares: too much inventory, or not enough". [46] For example, the firm's data mining efforts informed buyers that customers stock up on certain products in the days leading up to predicted hurricanes. Bumping

[42] Fichman, 2007.
[43] Friscia et al., 2009.
[44] Fishman, 2007.
[45] Joseph, 2009.
[46] Hays, 2004.

up supplies of batteries and bottled water is a no brainer, but the firm also learned that Pop-tarts sales spike seven fold before storms hit, and beer is the top pre-storm seller. This insight has lead to truckloads full of six packs and toaster pastries streaming into gulf states whenever word of a big storm surfaces[47].

Data mining also helps the firm predict things like how many cashiers are needed at a given store at various times of day throughout the year. And mined reports form the basis of the firm's weekly sales meetings, as well as executive strategy sessions.

Sharing Data, Keeping Secrets

While Wal-Mart is demanding of its suppliers, it also shares data with them, too. Data can help firms become more efficient so Wal-Mart can keep dropping prices, and data can help firms uncover patterns to help suppliers sell more. P&G's Gillette unit, for example, claims to have mined Wal-Mart data to develop promotions that can increase sales as much as 19 percent. More than 17,000 suppliers are given access to their products Wal-Mart performance across metrics that include daily sales, shipments, returns, purchase orders, invoices, claims and forecasts. And these suppliers collectively interrogate Wal-Mart data warehouses to the tune of 21 million queries a year. [48]

Many retailers pool their data by sharing it with information brokers like Information Resources and ACNielsen, in order to gain more collective insight. But not Wal-Mart. The firm stopped data sharing with these agencies years ago. It just didn't make sense to allow competitors access to such a potentially strategic asset.

The firm also custom-builds large portions of its systems to keep competitors off its trail. The Wal-Mart Data Center in McDonald County, Missouri was considered so secret that the county assessor was required to sign a non-disclosure statement before entering the site to determine property value.[49]

Challenges Abound

But despite success, challenges continue. While Wal-Mart grew dramatically throughout the 1990s, the firm's U.S. business has largely matured. And as a mature business it faces a problem not unlike the example of Microsoft discussed at the end of the Google chapter; Wal-Mart needs to find huge markets or dramatic cost savings in order to boost profits and continue to move its stock price higher.

Those low prices come at a price, and the firm's aggressiveness and shear size have increasingly made it a target for criticism. Wal-Mart has faced accusations of low wages and the firm remains a target of union activists. Others have accused the firm's suppliers of poor labor conditions. Suppliers that compete for Wal-Mart's business are often faced with a Catch 22. If

[47] Hays, 2004.
[48] Evans-Correia, 2006.
[49] McCoy, 2006.

they bypass Wal-Mart they miss out on the largest single chunk of world retail sales. But if they sell to Wal-Mart, the firm may demand prices so aggressively low that suppliers end up cannibalizing its own sales at other retailers. And the firm has battled local citizen groups that accuse Wal-Mart of ruining the market for mom and pop stores.

It's also important to know that big retailers don't survive forever. In the 1920s and 1930s, the A&P grocery chain once controlled 80 percent of U.S. grocery sales, at its peak operating five times the number of stores that Wal-Mart has today. But market conditions changed, and the government stepped in to draft anti-predatory pricing laws when it felt A&Ps parent was too aggressive.

For all of Wal-Mart's data brilliance, its warehouse wasn't able to provide external insights on the rise of Target and other up-market discounters. Historical data might offers little insight on how to adapt to more radical changes in the retail landscape. And another battle is brewing, as Tesco methodically attempts to take its globally honed expertise to U.S. shores. Data use is a vital tool, but not the only tool in management's strategic arsenal.

HARRAH'S: SOLID GOLD CRM FOR THE SERVICE SECTOR

Gary Loveman is a sort of management major trifecta. The CEO of Harrah's Entertainment is a former *operations* professor who has leveraged *information technology* to create what may be the most effective *marketing* organization in the service industry. If you ever needed an incentive to motivate you for cross-disciplinary thinking, Loveman provides it.

Harrah's has leveraged its data-powered prowess to move from an also-ran chain of casinos to become the largest gaming company by revenue. The firm operates some 53 casinos, employing more than 85,000 workers on five continents. Brands include Harrah's, Caesars Palace, Bally's, Horseshoe, and Paris Las Vegas. Under Loveman, Harrah's has aggressively swallowed competitors; the firm's $9.4 billion buyout of Caesars Entertainment being its largest deal to date.

Collecting Data

Data drives the firm. Harrah's collects customer data on just about everything you might do at their properties – gamble, eat, grab a drink, attend a show, stay in a room. The data's then used to track your preferences and to size up whether you're the kind of customer that's worth pursuing. Prove your worth, and the firm will surround you with top-tier service and develop a targeted marketing campaign to keep wooing you back[50].

The ace in the firm's data collection hole is its Total Rewards loyalty card system. Launched over a decade ago, the system is continually refined by an IT staff of 700, with an annual budget in excess of $100 million[51]. Total Rewards is an opt-in loyalty program, but customers consider

[50] Magnini et al., 2003.
[51] Swabey, 2007.

the incentives to be so good that the card is used by some 80 percent of Harrah's patrons, collecting data on over 44 million customers. [52]

Customers signing up for the card provide Harrah's with demographic information such as gender, age, and address. Visitors then present the card for various transactions. Slide it into a slot machine, show it to the restaurant hostess, present it to the parking valet, share your account number with a telephone reservation specialist – every contact point is an opportunity to collect data. Between 300,000 and 1 million customers come through Harrah's doors daily, adding to the firm's data stash and keeping that asset fresh[53].

Who Are the Most Valuable Customers?

All that data is heavily and relentlessly mined. Customer relationship management should include an assessment to determine which customers are worth having a relationship with. And because Harrah's has so much detailed historical data, the firm can make fairly accurate projections of *customer lifetime value* (*CLV*). CLV represents the present value of the likely future income stream generated by an individual purchaser[54]. Once you know this, you can get a sense of how much you should spend to keep that customer coming back. You can size them up next to their peer group, and if they fall below expectations, you can develop strategies to improve their spend.

The firm tracks over 90 demographic segments, and each responds differently to different marketing approaches. Identifying segments and figuring out how to deal with each involves an iterative model of mining the data to identify patterns, creating a hypothesis (customers in group X will respond to a free steak dinner; group Y will want $10 in casino chips), then testing that hypothesis against a control group, turning again to analytics to statistically verify the outcome.

The firm runs hundreds of these small, controlled experiments each year. Loveman says that when marketers suggest new initiatives "I ask, did we test it first? And if I find out that we just whole-hogged, went after something without testing it, I'll kill 'em. No matter how clever they think it is, we test it." [55] The former ops professor is known to often quote quality guru W. Edwards Deming, saying "In God we trust; all others must bring data".

When Harrah's began diving into the data, they uncovered patterns that defied the conventional wisdom in the gaming industry. Big money didn't come from European princes, Hong Kong shipping heirs, or the Ocean's 11 crowd – it came from locals. Less than 30 percent of customers who spent between $100 and $500 per visit accounted for over 80 percent of revenues and nearly 100 percent of profits[56].

---

[52] Wagner, 2008; and Haugsted, 2007
[53] Hoover, 2007.
[54] Knowledge@Wharton, 2003.
[55] Nickell, 2002.
[56] Swabey, 2007.

The data also showed that the firm's most important customers weren't the families that many Vegas competitors were trying to woo with 'Disneyland-style' theme casinos – it was Grandma! Harrah's focuses on customers 45 years and older. 20 somethings have no money. 30 somethings have kids and are too busy. To the pre-middle aged crowd, Loveman says, "God bless you, but we don't need you."[57]

Data Driven Service: Get Close (But Not Too Close) to Your Customers

The names for credit levels on the Total Rewards card convey increasing customer value – Gold, Diamond, and Platinum. Spend more money at Harrah's and you'll enjoy shorter lines, discounts, free items, and more. And if Harrah's systems determine you're a high-value customer, expect white-glove treatment. The firm will lavish you with attention, using technology to try to anticipate your every need. Customers notice the extra treatment that top-tier Total Rewards members receive, and actively work to improve their status.

To illustrate this, Loveman points to the obituary of an Ashville, NC woman who frequented a casino Harrah's operates on a nearby Cherokee Indian reservation. "Her obituary was published in the Asheville paper and indicated that at the time of her death, she had several grandchildren, she sang in the Baptist choir and she was *a holder of the Harrah's Diamond Total Rewards card*". Quipped Loveman, "When your loyalty card is listed in someone's obituary, I would maintain *you have traction*."[58]

The degree of customer service pushed through the system is astonishing. Upon check in, a Harrah's customer who enjoys fine dining may find their table is reserved, along with tickets for a show afterward. Others may get suggestions or special offers throughout their stay, pushed via text message to their mobile device[59]. The firm even tracks gamblers to see if they're suffering unusual losses, and Harrah's will dispatch service people to intervene with a feel-good offer: "Having a bad day? Here's a buffet coupon."[60]

The firm's CRM effort monitors any customer behavior changes. If a customer who usually spends a few hundred a month hasn't shown up in a while, the firm's systems trigger follow-up contact methods such as sending a letter with a promotion offer, or having a rep make a phone call inviting them back.[61]

Customers come back to Harrah's because they feel that those casinos treat them better than the competition. And Harrah's laser-like focus on service quality and customer satisfaction are embedded into its information systems and operational procedures. Employees are measured on metrics that include speed and friendliness, and are compensated based on guest satisfaction ratings. Hourly workers are notoriously difficult to motivate, they tend to be high-turnover, low-wage earners. But at Harrah's, incentive bonuses depend on an entire location's ratings. That

[57] Haugsted, 2007.
[58] Loveman, 2005.
[59] Wagner, 2008.
[60] Davenport and Harris, 2007.
[61] Loveman, 2003.

encourages strong performers to share tips to bring the new guy up to speed. The process effectively changed the corporate culture from an every-property-for-itself mentality to a collaborative, customer-focused enterprise[62].

While Harrah's is committed to learning how to make your customer experience better, the firm is also keenly sensitive to respecting consumer data. The firm has never sold or given away any of its bits to third parties. And the firm admits that some of its efforts to track customers have misfired, requiring special attention to find the sometimes subtitle line between helpful and 'too helpful'. For example, the firm's CIO has mentioned that customers found it "creepy and Big Brother-ish" when trying to greet customers by name and talk with them about their past business history at Harrah's, so the firm backed off.[63]

Innovation

Harrah's is constantly tinkering with new innovations that help it gather more data, and help push service quality and marketing program success. When the introduction of gaming in Pennsylvania threatened to divert lucrative New York City gamblers from Harrah's Atlantic City properties, the firm launched an interactive billboard in New York's Time Square, allowing passers-by to operate a virtual slot machine using text messages from their cell phones. Players dialing into the video billboard not only control the display, they receive text message offers promoting Harrah's sites in Atlantic City[64].

At Harrah's, tech experiments abound. RFID-enabled poker chips, and under-table RFID readers, allow pit bosses to track and rate game play far better than they could before. The firm is experimenting with using RFID-embedded bracelets for poolside purchases and Total Rewards tracking, for when customers aren't carrying their wallets. The firm has also incorporated drink ordering into gaming machines - why make customers get up to quench their thirst? A break in gambling is a halt in revenue.

The firm was also one of the first to sign on to use Microsoft's Surface technology – a sort of touch-screen and sensor-equipped tabletop. Customers at these tables can play bowling and group pinball games, and even pay for drinks using cards that the tables will automatically identify. Tech even helps Harrah's fight card counters and crooks, with facial recognition software scanning casino patrons to spot the bad guys[65].

Strategy

A walk around Vegas during Harrah's ascendency would find rivals with bigger, fancier casinos. Says Loveman, "We had to compete with the kind of place that God would build if he had the money… The only thing we had was data."[66]

---

[62] Magnini et al, 2003.
[63] Wagner, 2008.
[64] Economist Intelligence Unit, 2008.
[65] Lohr, 2007.
[66] Swabey, 2007.

That data advantage creates intelligence for a high-quality and highly personal customer experience. Data gives the firm a service differentiation edge. The loyalty program also represents a switching cost. And these assets combined to be leveraged across a firm that has gained so much scale that it's now the largest player in its industry, gaining the ability to cross-sell customers on a variety of properties – Vegas vacations, riverboat gambling, locally-focused reservation properties, and more.

Harrah's Chief Marketing Officer, David Norton points out that when Total Rewards started, Harrah's was earning about 36 cents on every dollar customers spent gaming – the rest went to competitors. A climb to 40 cents would be considered monstrous. By 2005 that number had climbed to 45 cents, making Harrah's the biggest monster in the industry[67]. Some of the firm's technology investments have paid back ten fold in just two years – bringing in hundreds of millions of dollars[68].

The firm's technology has been pretty tough for others to match, too. Harrah's holds several patents covering key business methods and technologies used in its systems. After being acquired by Harrah's, employees of Caesar's lamented that they had, for years, unsuccessfully attempted to replicate Harrah's systems without violating the firm's intellectual property[69].

Challenges

Harrah's efforts to gather data, extract information, and turn this into real profits is unparalleled, but it's not a cure-all. Broader events can often derail even the best strategy. Gaming is a discretionary spending item, and when the economy tanks, gambling one of the first things consumers will cut. Harrah's has not been immune to the world financial crisis, and experienced a loss in 2008.

Also note that if you look up Harrah's stock symbol you won't find it. The firm was *taken private* in January 2008 when buyout firms Apollo Management and TPG Capital paid $30.7 billion for all of the firm's shares. At that time Loveman signed a five year deal to remain on as CEO, and he's spoken positively about the benefits of being private – primarily that with the distraction of quarterly earnings off the table, he's been able to focus on the long term viability and health of the business[70].

But the firm also holds $24 billion in debt from expansion projects and the buyout, all at a time when economic conditions have not been favorable to leveraged firms[71]. A brilliantly successful firm leveraging best-in-class customer relationship management in now in a position many consider risky due to debt assumed as part of an overly optimistic buyout occurring at precisely

[67] Lundquist, 2005.
[68] Swabey, 2007.
[69] Hoover, 2007.
[70] Knightly, 2009.
[71] Lattman, 2009.

the time the economy went into a terrible funk. Harrah's awesome risk-reducing, profit pushing analytics failed to offer any insight on the wisdom (or risk) in the debt and private equity deals.

• • • • •

About This Work

The goal of this project is to have an impact. At my university, we've bucked the national trend, tripling our Information Systems majors in the three years since we've adopted a business-focused IS teaching approach, and I'm delighted to share this content with you. I hope that Flatworld's free online copies and low-cost print versions encourage wide adoption of this material, and I hope that you and your students enjoy it. Please tell others, and thanks!

Comments & feedback are most welcome! Contact Info:

- E-Mail: john.gallaugher@bc.edu
- Draft Chapters, Cases, Slides, and Podcasts: http://gallaugher.com/chapters
- For updates, supporting articles, and commentary, sign up for the Week in Geek at: http://www.gallaugher.com
- And follow on Twitter at http://twitter.com/gallaugher

About the Author

John Gallaugher is a member of the Dept. of Information Systems in Boston College's Carroll School of Management. Prof. Gallaugher teaches courses and conducts research at the intersection of technology and strategy. An award-winning teacher, he leads the School's TechTrek programs, co-leads the Asian field study program, and has consulted to and taught executive seminars for several organizations including Accenture, Alcoa, Brattle Group, ING Group, Patni Computer Systems, Staples, State Street, and the U.S. Information Agency.

REFERENCES

Appleton, R., "Less Independent Doctors Could Mean More Medical Mistakes", *InjuryBoard.com*, June 14, 2009.

Baker, S., "Math Will Rock Your World", *BusinessWeek*, Jan. 23, 2006.

Baker, S., *The Numerati*, Houghton Mifflin Harcourt, New York, 2008.

Baker, S., "Data Mining Moves to Human Resources", *BusinessWeek*, March 12, 2009.

Byrnes, N., "Coming of Age at Capital One", *BusinessWeek*, Feb. 16, 2004.

Braddock, R., "Lessons of Internet Marketing from Fresh Direct", *The Wall Street Journal*, May 11, 2009.

Capell, K., "Tesco: 'Wal-Mart's Worst Nightmare'", *BusinessWeek*, Dec. 29, 2008.

Capell, K., "Tesco Hits Record Profit, But Lags in U.S.", *BusinessWeek*, April 21, 2009.

Conry-Murray, A., "The Pain of e-Discovery", *InformationWeek*, June 1, 2009

Computewrold Technology Briefings, "Intelligent Users Use Business Intelligence", 2009.

Coy, P., "He Who Mines Data May Strike Fool's Gold", *BusinessWeek*, June 16, 1997.

Davenport, T., and Harris, J., *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, Boston, 2007.

Davenport, T., and Harris, J., "Competing with Multichannel Marketing Analytics", *Advertising Age*, April 2, 2007b.

Esfahani, E., "7-Eleven Gets Sophisticated", Business 2.0, Jan. 1, 2005.

Evans-Correia, K., "Dillman replaced as Wal-Mart CIO", *SearchCIO*, April 6, 2006.

Dragoon, A., "Business Intelligence Gets Smart(er)", *CIO*, Sept.. 15, 2003.

Economist Intelligence Unit, "The Global CMO", Sept. 2008.

Fichman, C., "The Wal-Mart You Don't Know", *Fast Company*, Dec. 19, 2007.

Friscia, T., O'Marah,K., Hofman, D., Souza , J., "The AMR Research Supply Chain Top 25 for 2009", *AMR Research*, May 28, 2009.

Gefter, A., and Simonite, T., “What The Data Miners are Digging Up About You”, *CNet*, Dec. 1, 2008.

Gewirtz, D., “What if someone stole the Library of Congress?”, *CNN.com/AC360*, May 25, 2009.

Greenberg, A., “Companies that Profit From Your Data”, *Forbes*, May 14, 2008.

Halamka, J., “IT Spending: When Less is More”, *BusinessWeek*, March 2, 2009.

Hansell, S., “How Wall Street Lied to Its Computers”, The New York Times, Sept. 18, 2008.

Harvey, M., “Probe into how Google mix-up caused $1 billion run on United”, *Times Online*, Sept. 12, 2008.

Haugsted, L., “Better Take Care of Big Spenders; Harrah’s Chief Offers Advice to Cablers”, *Multichannel News*, July 30, 2007.

Havenstein, H., “HP nabs Wal-Mart as data warehousing customer”, ComputerWorld, Aug. 1, 2007.

Hays, C., “What Wal-Mart Knows About Customer Habits”, *The New York Times*, Nov. 14, 2004.

Helm, B., “Getting Inside the Customer's Mind”, *BusinessWeek*, Sept. 11, 2008.

Hoover, N., “Chief of the Year: Harrah’s CIO Tim Stanley”, *Information Week Research and Reports*, 2007.

Joseph, D., “Supermarket Strategies: What's New at the Grocer”, *BusinessWeek*, June 8, 2009.

Kahn, J., “It’s Alive”, Wired, March, 2002.

Kemp, Ted, “PrivacyFlap Cuts The Click in DoubleClick”, *InternetWeek*, Dec. 18, 2000.

King, R., “Business Intelligence Software’s Time is Now”, *BusinessWeek*, March 2, 2009.

King, R., “Intelligence Software for Business”, BusinessWeek Podcast, Feb. 27, 2009.

King, J., “Can Web 2.0 Save BI?” *ComputerWorld*, Sept. 1, 2008.

Klebnikov, P., “The Resurrection of NCR”, *Forbes*, July 9, 2001.

Knowledge@Wharton, “Which Customers Are Worth Keeping and Which Ones Aren't? Managerial Uses of CLV”, July 30, 2003.

Koch, C., “Who’s Mining the Store?”, *CIO*, May 15, 2005.

Lattman, P., “A Buyout-Shop Breather”, *The Wall Street Journal*, May 30, 2009.

Levinson, M., "Harrah's Knows What You Did Last Night," Darwin Magazine, May 2001.

Lundquist, E., “Harrah’s Bets Big on IT”, *eWeek*, July 20, 2005.

Lohr, S., “Reaping Results: Data-Mining Goes Mainstream”, *The New York Times*, May 20, 2007.

Lowenstein, M., “Tesco: A Retail Customer Divisibility Champion”, *CustomerThink*, Oct. 20, 2002.

Loveman, G., “Diamonds in the Data Mine”, *Harvard Business Review*, May 2003.

Loveman, G., Speech and Comments, Chief Executive Club of Boston College, January 2005.

Magnini, V., Honeycutt, E., Hodge, S,. “Data mining for hotel firms: use and limitations”, *Cornell Hotel & Restaurant Administration Quarterly*, April 2003.

McCoy, M., “Wal-Mart's data center remains mystery”, *The Joplin Globe*, May 28, 2006.

McCullagh, D., “Q&A: Electronic Health Records And You”, *CNet/CBSNews.com*, May 19, 2009.

McKay, L, “Decisions, Decisions”, *CRM Magazine*, May 1, 2009.

Mearian, L., “IT organizations will need to transform their existing relationships with business units”, *ComputerWorld*, March 18, 2008.

Milligan, S., “Business warms to Democratic leaders”, *The Boston Globe*, May 28, 2009.

Mithchell, R., “Why You Should Be Worried About Your Privacy on the Web”, *Computerworld*, May 11, 2009.

Mulcahy, R., “ABC: An Introduction to Business Intelligence”, *CIO*, March 6, 2007.

Newswire, “Web sites pull back from Do0ubleClick amid privacy controversy”, *Associated Press*, March 12, 2000.

Nickell, J., "Welcome to Harrah's," Business 2.0, April 2002.

Obama, B., President’s Speech to the American Medical Association, June 15, 2009.

Port, O., “Thinking Machines”, *BusinessWeek*, Aug. 7, 2000.

Ranger, S., “How Firms Use Business Intelligence”, *BusinessWeek*, May 24, 2007.

Rigby, D., and Ledingham, D., “CRM Done Right”, *Harvard Business Review*, Nov. 2004.

Searcy, D., “Employers Watching Workers Online Spurs Privacy Debate, *The Wall Street Journal*, April 23, 2009.

Soat, J., “P&G's CIO Puts IT At Users' Service”, *InformationWeek*, Dec. 15, 2007.

Swabey, P., “Nothing Left to Chance”, *Information Age*, January 18, 2007.

Swarns, R., “Senator? Terrorist? A Watch List Stops Kennedy at Airport “, *The New York Times*, Aug. 20, 2004.

Wagner, M., “Harrah's Places Its Bet On IT”, *InformationWeek*, Sept. 16, 2008.

Wahba, P., “Buffeted "quants" are still in demand”, *Reuters*, Dec. 22, 2008.

Wailgum, C., “How Wal-Mart Lost Its Technology Edge”, *CIO*, Oct. 4, 2007.

Wailgum, T., “Business Intelligence: A Technology Category in Tumult”, *CIO*, March 24, 2008.

Warmenhoven, D., “Protect Me, Protect My Data”, *BusinessWeek*, June 8, 2006.

Zhang, J., “Recession Likely to Boost Government Outlays on Health Care”, *The Wall Street Journal*, Feb. 24, 2009.